



096419

73-0001

~~3,03.02~~



LM096419

1  
20

Coordination Of Computerized  
Information Systems Reporting  
On Active Research Efforts B-115398

Office of Management and Budget

**UNITED STATES  
GENERAL ACCOUNTING OFFICE**

~~701475~~ 096419

MARCH 29, 1973



UNITED STATES GENERAL ACCOUNTING OFFICE

WASHINGTON, D.C. 20548

PROCUREMENT AND SYSTEMS  
ACQUISITION DIVISION

B-115398

The Honorable Roy L. Ash, Director  
Office of Management and Budget 27

Dear Mr. Ash:

The General Accounting Office, in its continuing study of the management of research and development activities, reviewed the status of coordinating computerized information systems used by agencies for reporting on active research efforts. We are bringing the results of our review to your attention for use in considering the future role of the Smithsonian Science Information Exchange.

In a report entitled "Effectiveness of Smithsonian Science Information Exchange Hampered by Lack of Complete, Current Research Information" (B-175102, Mar. 1, 1972), we recommended that the Director, Office of Management and Budget (OMB), examine the need for the Exchange as part of OMB's responsibility for fostering coordination of Federal programs. We recommended also that, if it decided the Exchange should be continued, OMB require all Federal agencies to submit pertinent information to the Exchange promptly so that the Exchange can serve its intended purpose effectively.

In commenting on our report, OMB said that further study of the Exchange would be appropriate and that the Smithsonian Institution had agreed to contract for such a study. OMB said also that it would closely review the study at each stage and would decide the future of the Exchange on the basis of the study results.

On July 31, 1972, the Smithsonian Institution contracted with Research, Planning and Management Services For the 70's to study the Exchange and its role in research management.

Our staff paper on "Progress in Achieving a Coordinated Information System on Active Research," prepared in March 1972, was discussed at a meeting on March 23, 1972, with representatives of many agencies. We hoped to obtain their views on the advantages or disadvantages of a coordinated

information system for active research. In the paper we proposed that present and future agency data banks be connected to form a Government-wide research information network, and we invited the agencies to comment on the material.

Reactions to our proposal varied; however, there was general agreement that improved communication of active research information was desirable. Most agencies felt that there were too many unknown factors concerning the feasibility of establishing a network system for transferring data. A major concern was that there were many different levels and types of potential users of active research data whose various needs for information were not well known. Several agencies said that developing a Government-wide system for exchanging research information would be extremely complex and would require extensive study and careful evaluation.

Many agencies suggested that improving the existing Exchange would be a reasonable alternative to establishing a network system and would be a logical step toward improved coupling of research information. For example, the Department of Health, Education, and Welfare said that, if the Exchange were developed to its full potential, it would more adequately meet the agencies' needs. Because a study of the Exchange is underway, we believe that it would be appropriate at this time to make available to OMB the additional information we obtained.

#### BACKGROUND

Federal funds for research and development have increased from about \$1 billion in fiscal year 1950 to over \$17 billion in 1972. According to the National Science Foundation, 33 agencies budgeted about \$6.6 billion in fiscal year 1972 for research defined as "a systematic, intensive study directed toward fuller scientific knowledge or understanding of the subject studied." One matter of concern in improving the Government's ability to manage a large and diverse science program has been the need to improve the flow of information to the top levels of Government and to coordinate large and often overlapping research activities among agencies. Both the executive branch and the Congress are concerned with how to communicate information about current research efforts.

Major studies since 1945, sponsored or conducted by the Executive Office of the President and the congressional committees, recognize the importance of making information on active research projects widely available and the need for improved methods of communicating such information.

Although the studies differed as to the means of achieving this goal, they generally recognized that some type of automated inventory system or project index containing data on active research projects was needed if managers, scientists, and others were to have basic data on who was doing what research and where.

QUESTIONNAIRE RESULTS INDICATE  
CURRENT SYSTEMS PROVIDE USEFUL DATA

To determine the need for active research project data and the use of existing agency systems by various levels of management within the Federal Government, we surveyed by questionnaire a random sample of the approximately 262,000 Federal professional scientific, technical, and health personnel.

About 51 percent of the respondents needed information, at least once a year, on research in progress either in their agencies or in other agencies. Of these, 38 percent found the existing data banks useful sources.

The questionnaire identified various uses of agency data banks. We found that the data banks referred to in our questionnaire were used primarily in support of technical tasks and, to a lesser degree, for management purposes.

About 99 percent of the respondents using the specified agency data banks rated the material received as satisfactory or better. Although many users thought the data was incomplete, they believed it sufficient for their purposes. The most frequently identified general benefits of using the data banks were: searching time was lessened, duplication was minimized, and communication among researchers was improved.

We asked the respondents to estimate any savings attributable to using the current data banks in the last 12 months. From the savings estimated by the users, we projected that the total amount of funds that could have been saved or redirected as a result of information received from these data banks was between \$9 million and \$42 million.

Fifteen percent of the respondents said that they had discovered unplanned duplication of research work through various sources, and half of them had discontinued their research projects upon learning of the duplication.

The questionnaire results tend to support the general hypothesis that scientists and research managers keep informed

about current research through informal contacts and attendance at scientific meetings.

An analysis of the questionnaire results is included as the appendix.

### CONCLUSION

Although the problem of improving the flow of data on active research is both massive and complicated, without ready solutions, we believe that the desire and willingness of some agencies to work toward an improved Exchange could provide an acceptable alternative in achieving an improved information service.

If the contract study shows that the Exchange is needed and it is decided to continue it, agencies should consider using the Exchange to the maximum in meeting the needs of the Government for information on active research projects.

OMB and the Department of the Treasury are developing a standard budgetary and fiscal information system required by title II of the Legislative Reorganization Act of 1970. In a report entitled "Budgetary and Fiscal Information Needs of the Congress" (B-115398, Nov. 10, 1972), we described our assessment of the information needs identified by Committees and Members of the Congress. For example, the House Committee on Science and Astronautics needs to easily obtain basic financial information on individual agency and overall Federal research and development activities. Because many agencies include funding information in the research project data submitted to the Exchange, an improved Exchange could possibly meet the needs of this Committee for information on active research projects. H 03460

We hope the results of our study are useful to your Office and to the National Science Foundation in carrying out their leadership roles relative to science activities. Copies of this report will be sent to the Director, National Science Foundation, and the other agencies that participated in the meeting of March 23, 1972.

Sincerely yours,

*for* 

Director

ANALYSIS OF QUESTIONNAIRE RESULTS  
ON THE  
GOVERNMENT-WIDE RESEARCH INFORMATION NEEDS  
OF SCIENTIFIC AND TECHNICAL PERSONNEL  
WITHIN THE FEDERAL GOVERNMENT

	<u>Page</u>
I Characteristics of the sample	1
II Current information sources	3
III Utilization of data banks for ongoing research information	6
IV Future information sources	9

I. CHARACTERISTICS OF THE SAMPLE

From an estimated total of 262,000 Federal professional scientific, technical, and health personnel, we sent questionnaires to a random sample of 619. Of these, 579 or almost 94 percent returned the questionnaires, with 537 responding to the individual questions.

The results are statistically representative of the 262,000 personnel at the 95 percent confidence level. Thus, there is only a 1 in 20 chance that the percentage derived from the sample would differ by more than the sampling errors (S.E.) shown.

Specified characteristics of personnel  
included in the sample (note 1)

Classified by field of science:

	<u>Percent</u>	<u>S.E. (±)</u>
Physical	16	3.0
Mathematics and statistics	4	1.6
Biological	10	2.5
Social	4	1.6
Engineering	34	3.9
Health	22	3.4
Other	3	1.4
No response	<u>7</u>	2.1
	100	

Classified by level of education:

	<u>Percent</u>	<u>S.E. (±)</u>
High school	9	2.3
Bachelor's degree	47	4.1
Master's degree	20	3.3
Doctor's degree	10	2.5
Medical doctor	5	1.8
Other	2	1.0
No response	<u>7</u>	2.1
	100	

<sup>1</sup>Based on the 579 returned questionnaires

APPENDIX

Classified by civil service grade levels or their equivalents:

	<u>Percent</u>	<u>S.E. (±)</u>
GS-12 and below	57	4.0
GS-13, 14	26	3.6
GS-15 and above	9	2.4
No response	7	2.1

Classified by primary work functions:

	<u>Percent</u>	<u>S.E. (±)</u>
Clinical practice, counseling, and ancillary medical services	15	3.0
Scientific and technical research	14	2.8
Management (other than direct supervision-included in the other functions)	8	2.2
Development	8	2.2
Design	6	2.0
Planning	6	1.9
Regulatory enforcement and licensing	4	1.6
Test and evaluation	4	1.5
Installation, operations, and maintenance	3	1.5
Data collection, processing, and analysis	3	1.5
Natural resource operations	3	1.3
Technical assistance and consulting	2	1.3
Teaching and training	2	1.2
Construction	2	1.1
Standards and specifications	2	1.1
Research contract and grant administration	1	.9
Production	1	.8
Scientific and technical information	1	.7
Other	8	2.2
No response	<u>7</u>	2.1
	100	

Classified by geographic location:

	<u>Percent</u>	<u>S.E. (±)</u>
Washington, D.C. area	15	2.9
Field	78	3.4
No response	<u>7</u>	2.1
	100	

Classified by those who have needs for information, at least once a year, about current research in progress within or outside their own agencies:

<u>By field of science</u>	<u>% within own agency</u>	<u>% outside own agency</u>	<u>% within or outside own agency</u>
Physical	71	69	74
Mathematics and statistics	45	41	45
Biological	63	49	66
Social	74	52	83
Engineering	48	44	53
Health	33	26	37
Other	47	47	53
Total (including "no responses")	47	42	51 <sup>a</sup>
<u>By grade level, or equivalent</u>			
GS-12 and below	48	40	52
GS-13, 14	54	50	58
GS-15 and above	67	59	70

## II. CURRENT INFORMATION SOURCES

Current information sources utilized by persons included in the sample, classified by source of information, and use of information. This table is based on the number of respondents who answered questions 6, 7, and 8. The number of respondents is shown at column head.

<u>Source of information</u>	<u>To learn about research in progress</u>		<u>For managing research activities</u>
	<u>Within own agency</u> (base = 290)	<u>In other agencies</u> (base = 277)	<u>(base = 223)</u>
Data banks with ongoing research information:			
Percent of the respondents using information from at least one of the following data banks	33% <sup>a</sup>	32%	29%
Smithsonian Science Information Exchange (SSIE)	9	5	6
USDA Current Research Information System (CRIS)	8	9	7
Defense Documentation Center, Work Unit Information System (DDC/WUIS)	13	14	11
Others	11	9	8
Notification of research work about to start	14	8	14
Preconvention published proceedings	11	17	7
Copies of convention presentation	19	21	13

Source of information (continued)

Published convention program	15%	20%	14%
Attendance at local scientific and technical meetings	34	37	30
Attendance at State, regional, or national conventions	30	35	30
Informal personal contacts	71	59	54
Theses	6	6	4
Manuscripts	8	7	9
Formal progress reports	32	18	27
Technical reports	48	44	48
Preprints prior to submission to journal	9	9	9
Information analysis center	0	0	1
Other	8	8	5

Note: Percentages will not add to 100% because respondents could check more than one source.

<sup>a</sup>38% of those who indicated a need for current research information from either within or outside their own agencies found data banks to be useful sources.

As demonstrated by these results, data banks for information on current research are considered to be high among the most useful for learning about current research, and, in our opinion, a valuable supplement to the other sources. We believe that, although informal means seem to be the most prevalent now, the use of data banks will grow in their complementary relationship with the other more traditional sources as the reliability of data banks for providing complete and current information is increased.

Sources of information used for discovering research duplications

BEST DOCUMENT AVAILABLE

Personnel who have discovered, within the past 2 years, at least one instance where they had started research that had already been done--classified by source of information:

	Est. number	S.E. ( $\pm$ )	Percent	S.E. ( $\pm$ )
Total	39,400	7,600	15.0	2.9
<u>Source of information</u>				
Informal personal contacts	21,300	5,800	8.1	2.2
Technical reports	10,900	4,300	4.2	1.6
Attendance at State, regional or national conventions	8,100	3,700	3.1	1.4
Attendance at local scientific and technical meetings	6,800	3,400	2.6	1.3
Data banks	5,900	3,200	2.3	1.2
Other sources (1.6% or less each source)				

Note: Figures do not add to the totals because respondents could check more than one source.

Personnel dropping projects for which duplication was discovered, by source of information:

	<u>Est. number</u>	<u>S.E. (±)</u>	<u>Percent</u>	<u>S.E. (±)</u>
Total	19,900	5,700	7.6 <sup>a</sup>	2.2
<u>Source of information</u>				
Informal personal contacts	10,000	4,100	3.8	1.6
Technical reports	5,400	3,000	2.1	1.2
Attendance at State, regional or national conventions	4,500	2,800	1.7	1.1
Attendance at local scientific and technical meetings	3,600	2,500	1.4	1.0
Data banks	3,600	2,500	1.4	1.0
Other sources (1% or less each source)				

<sup>a</sup>50% of those who discovered duplication.

Note: Figures do not add to the totals because respondents could check more than one source.

Personnel discovering duplication, classified by time elapsed between discovery of the duplication and the publication of its results, and by location of duplication:

<u>Time elapsed, months</u>	<u>Percent finding duplication</u>	
	<u>within own agency</u>	<u>from outside own agency</u>
1-6	30	19
7-12	36	35
13-18	16	29
19-24	7	6
Longer	<u>11</u>	<u>11</u>
	100	100

In summary, it appears that unplanned duplication of research effort may occur quite frequently. The discovery of duplication may be quickened by making information within the data banks more complete and current.

### III. UTILIZATION OF DATA BANKS FOR ONGOING RESEARCH INFORMATION

Personnel who were aware of the Federal data banks which contain information on ongoing research:

	<u>Est. number</u>	<u>Percent</u>	<u>S.E. (±)</u>
Smithsonian Science Information Exchange (SSIE)	43,400	17	3.0
USDA Current Research Information Service (CRIS)	38,500	15	2.9
Defense Documentation Center, Work Unit Information System (DDC/WUIS)	53,400	20	3.3
Others	29,900	11	2.6

#### Utilization of each data bank

Classified by frequency of use:

	<u>Estimated number of users</u>	<u>S.E. (±)</u>	<u>Percent</u>	<u>S.E. (±)</u>	<u>Average times used by each user per year</u>
SSIE	6,800	3,400	2.6	1.3	3.3
CRIS	4,500	2,800	1.7	1.1	3.4
DDC	19,900	5,700	7.6	2.2	4.1
Others	<u>7,200</u>	3,500	2.8	1.3	3.7
Total	38,400	7,600	14.7	2.9	3.8

Classified by purpose of use:

	<u>Overall percent</u>	<u>Number of responses</u>			
		<u>SSIE</u>	<u>CRIS</u>	<u>DDC</u>	<u>Others</u>
Management purposes	15	3	2	4	3
Support of technical tasks	57	8	2	29	5
An equal amount of each	<u>28</u>	<u>2</u>	<u>4</u>	<u>11</u>	<u>5</u>
	100	13	8	44	13

Classified by evaluation of the data:

	<u>Overall percent</u>	<u>Number of responses</u>			
		<u>SSIE</u>	<u>CRIS</u>	<u>DDC</u>	<u>Others</u>
Extremely valuable	13	2	2	3	3
Valuable	50	4	2	22	10
Satisfactory	36	6	2	17	2
Unsatisfactory	<u>1</u>	<u>0</u>	<u>0</u>	<u>1</u>	<u>0</u>
	100	12	6	43	15

## Classified by completeness of data:

	Overall percent	Number of responses			
		SSIE	CRIS	DDC	Others
Complete	22	2	2	7	6
Incomplete but highly sufficient	34	4	4	13	5
Incomplete but sufficient	33	2	1	17	5
Incomplete and inadequate	<u>11</u>	<u>2</u>	<u>0</u>	<u>6</u>	<u>0</u>
	100	10	7	43	16

## Classified by the average number of days between requesting and receiving the data:

	Overall average	SSIE	CRIS	DDC	Others
	18 days	18	13	21	10

## Classified by specific benefits received:

	Overall percent	Number of responses			
		SSIE	CRIS	DDC	Others
Make the process of search less time-consuming	69	8	5	27	12
Minimize duplication	65	8	4	29	8
Aid communication between researchers	55	7	3	23	8
Identify levels of technical effort in various scienti- fic and technical fields	47	6	2	20	7
Become aware of negative results	41	6	1	19	5
Identify newly sponsored work to compare with pro- ject proposals	37	6	3	13	6
Identify specialists and information in unrelated fields for background on an interdisciplinary task	37	5	3	15	5

APPENDIX

	Overall percent	Number of responses			
		SSIE	CRIS	DDC	Others
Identify balance of basic and applied research	28	5	2	7	7
Identify personnel to receive research results	28	3	2	14	2
Salvage information from canceled projects	27	4	1	12	3
Identify distribution of research among Government, labs, industry, universi- ties, nonprofit organiza- tions, etc.	24	3	3	7	5
Determine that the most important problems are receiving adequate attention	24	3	3	7	5
Identify division of support for R&D in different areas as between Federal and non- Federal sponsorship	17	3	2	4	4
Identify location of unique scientific equipment	16	3	2	6	1
Match problems with workers and facilities	15	4	3	3	1
Prepare budget data	9	0	1	4	2
Identify availability of per- sonnel for research projects suffering from insufficient manpower	5	1	2	1	0
Total respondents		12	7	41	15

The benefits experienced by these systems users are close to the ranking of benefits desired by all persons--including noncurrent systems users--who indicated their need for a Government-wide coordinated information system, as shown on page 11. This appears to support the feasibility of a future Government-wide system providing the desired benefits.

BEST COPY AVAILABLE

Estimated dollar savings resulting from the use of data banks:

	<u>Total</u>	<u>Number of responses</u>			
		<u>SSIE</u>	<u>CRIS</u>	<u>DDC</u>	<u>Others</u>
\$0	13	1	1	8	3
1 - 500	16	3	0	9	4
501 - 1,000	10	2	2	4	2
1,001 - 2,000	4	0	0	4	0
2,001 - 5,000	2	0	0	1	1
5,001 - 10,000	4	0	1	3	0
Over 10,000	4	<u>1</u>	<u>0</u>	<u>2</u>	<u>1</u>
Total		7	4	31	11

This section demonstrates the significant amount of funds that are claimed to be saved or redirected as a result of information received from these data banks. We made a few conservative assumptions about these results and then arrived at the following estimates of total savings:

Total estimated savings experienced by the systems users within the sample

\$ 60,000

Total projected estimated savings experienced by all Federal professional scientific, technical and health personnel users

25,400,000  $\pm$  16,700,000

These estimates were based on the following conservative assumptions and exclusion:

- assume each user saved the lowest of the dollar range claimed.
- assume the dollar savings represent total savings attributable to each user instead of to each time the system was used.
- exclude results from the "Other" system, as some of these systems are not truly data banks for ongoing research information.

#### IV. FUTURE INFORMATION SOURCES

We included in the questionnaire the following description of a potential information system and then asked whether the responders thought such a system could be of use to them:

"There has been considerable concern expressed about the need for development of methods through which today's technical personnel can be kept better aware of research work of others.

## APPENDIX

"It has been proposed that a Government-wide information system for ongoing research and technology efforts be established which would have automated access to summary information on all unclassified ongoing work sponsored by the Federal Government. This information would basically consist of summary descriptions of individual projects or 'Work Units' (the natural unit into which science and technology work is normally divided for purposes of local administration). In addition, with each description there would be various identifiers, such as keywords, responsible agency, contractor, responsible personnel, dates, funds, and procurement methods. In other words, who is doing what work when and where. This data would be included in the system within 15 days of each research project's starting date.

"Typical questions that could be answered by this system include:

- Current efforts on development of high strength steels.
- All research related to pesticides sorted according to species of pests, i.e., insecticides, rodenticides, fish poison, and so on.
- All research on immigration to urban areas and ensuing problems.
- All defense contracts or grants supporting research in foreign universities.

"Assuming the system as described above would become fully operative (it is now partially so) and the ease of submitting questions by you to the system would entail nothing more than sending most questions directly to an online system terminal which would supply the answers within minutes: Do you think the proposed system could be of use to you?"

Sixty six percent of the 579 respondents answered yes, 20 percent no, and 13 percent had no response. Projecting this 66 percent to all Federal professional scientific, technical, and health personnel, results in an estimated number of about 173,000 potential users of the proposed Government-wide system (over four times the estimated number of current systems users (see page 6)).

The following table shows how these responses for the Government-wide system compare to responses on the needs for ongoing research information from outside their agencies (see page 3) and to responses on the utilization of existing data banks (see page 6). The responses vary somewhat according to the primary work functions of the respondents, but there appear to be no significant differences based on grade levels. As expected, scientific and technical researchers express the greatest needs.

## APPENDIX

	Desire a		Need information		Use existing	
	Government-wide		external to		data banks	
	system		own agency			
	Percent	S.E. (±)	Percent	S.E. (±)	Percent	S.E. (±)
Total	66	3.8	42	4.0	13	2.7
By primary work function						
--with sample sizes over 40						
Scientific	91	6.3	87	7.3	29	10.0
Development	86	10.1	57	14.6	32	13.8
Management	72	12.8	51	14.3	11	8.8
Clinical practices	54	10.3	20	8.3	0	- a
--all others	60	5.4	33	5.2	10	3.3
By grade level--or equivalent						
GS-12 and below	70	4.9	40	5.3	12	3.5
GS-13, 14	76	6.9	50	8.0	15	5.7
GS-15 and above	63	12.9	59	13.1	17	10.0

<sup>a</sup>95% confidence limits = 0% to 4.1%

Specific aids the respondents want from the proposed Government-wide system:

Abbreviated description of identified benefits	Sample results:		Projected estimate to all Federal professional scientific, technical, and health personnel (approximately 262,000) from which the sample was drawn		
	Percent	S.E. (±)	GS-12		Total
			and below	and above	
Aid communication between those engaged in parallel work	48	4.1	74,200	51,100	125,300
Make the process of search less time-consuming	48	4.1	78,700	46,600	125,300
Minimize duplication	40	4.0	62,000	43,400	105,400
Become aware of negative results	30	3.7	46,200	32,100	78,300
Identify levels of tech- nical effort in various scientific and technical fields	28	3.7	43,000	30,800	73,800
Identified newly sponsored work to compare with project proposals	26	3.6	33,900	33,000	66,900
Aid in identifying infor- mation for interdisciplin- ary research task work	25	3.5	34,800	29,900	64,700

(continued)

## APPENDIX

<u>Abbreviated description of identified benefits</u>	<u>Sample results:</u>		<u>Projected estimate to all Federal professional scientific, technical, and health personnel (approximately 262,000) from which the sample was drawn</u>		
	<u>Percent</u>	<u>S.E.(±)</u>	<u>GS-12 and below</u>	<u>GS-13 and above</u>	<u>Total</u>
	Salvage information from canceled projects	24	3.5	37,100	24,900
Determine that the most important problems are receiving adequate attention	20	3.3	33,900	18,600	52,500
Identify personnel to which significant find- ings should be disseminated	18	3.1	31,200	16,300	47,500
Identify location of unique scientific apparatus	16	3.0	26,200	15,800	42,000
Identify distribution of research between Govern- ment laboratories, indus- try, universities, etc.	14	2.8	23,100	13,600	36,600
Identify division of sup- port for R&D in different areas as between Federal and non-Federal sponsor- ship	12	2.7	20,400	11,800	32,100
Match important problems with scientific personnel and facilities	10	2.5	18,100	8,600	26,700
Prepare budget data	9	2.4	13,600	11,300	24,900
Identify balance of basic and applied research	9	2.4	16,700	7,200	24,000
Identify availability of personnel for research projects suffering from insufficient manpower	8	2.2	9,500	10,400	19,900

BEST DOCUMENT AVAILABLE